

DOCUMENT RESUME

ED 144 999

TM 006 605

AUTHOR Prestwood, J. Stephen; Weiss, David J.  
TITLE Accuracy of Perceived Test-Item Difficulties.  
Research Report No. 77-3.  
INSTITUTION Minnesota Univ., Minneapolis. Dept. of Psychology.  
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel  
and Training Research Programs Office.  
REPORT NO NRP-RR-77-3  
PUB DATE May 77  
CONTRACT N00014-76-C-0243  
NOTE 38p.

EDRS PRICE MF-\$0.83 HC-\$2.46 Plus Postage.  
DESCRIPTORS Ability; \*Complexity Level; Higher Education; Item  
Analysis; \*Perception; Student Reaction; Test  
Construction; \*Test Items; Undergraduate Students;  
\*Verbal Ability

ABSTRACT

The accuracy with which testees perceived the difficulty of ability-test items was investigated by administering two 41-item conventional tests on verbal ability. High- and low-ability groups responded to test items by choosing the correct alternative and then rating each item's difficulty relative to their levels of ability. Least-squares estimates of item difficulty correlated highly with proportion-correct and latent trait estimates of item difficulty based on a norming sample. Least-squares estimates of testee ability, which were based solely on the difficulty perceptions of the testees, correlated significantly with number-correct and maximum-likelihood ability scores based on the testees' responses to the items. These results show that item-difficulty often used in test construction, and that as testee ability level increased, the items were perceived as being relatively less difficult. The relationship between a testee's ability and his/her perception of an individual item's relative difficulty appeared to be weak. Of major importance was the finding that items, which were appropriate in difficulty levels from a psychometric standpoint were perceived by the testees as being too difficult for their ability levels. (Author/DSE)

\*\*\*\*\*  
\* Documents acquired by ERIC include many informal unpublished \*  
\* materials not available from other sources. ERIC makes every effort \*  
\* to obtain the best copy available. Nevertheless, items of marginal \*  
\* reproducibility are often encountered and this affects the quality \*  
\* of the microfiche and hardcopy reproductions ERIC makes available \*  
\* via the ERIC Document Reproduction Service (EDRS). EDRS is not \*  
\* responsible for the quality of the original document. Reproductions \*  
\* supplied by EDRS are the best that can be made from the original. \*  
\*\*\*\*\*

THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIGIN-  
ATING IT. POINTS OF VIEW OR OPINIONS  
STATED DO NOT NECESSARILY REPRESENT  
OFFICIAL NATIONAL INSTITUTE OF  
EDUCATION POSITION OR POLICY.

# ACCURACY OF PERCEIVED TEST-ITEM DIFFICULTIES

J. Stephen Prestwood  
and  
David J. Weiss

RESEARCH REPORT 77-3

MAY 1977

PSYCHOMETRIC METHODS PROGRAM  
DEPARTMENT OF PSYCHOLOGY  
UNIVERSITY OF MINNESOTA  
MINNEAPOLIS, MN 55455

Prepared under contract No. N00014-76-C-0243, NR150-382  
with the Personnel and Training Research Programs  
Psychological Sciences Division  
Office of Naval Research

Approved for public release; distribution unlimited.  
Reproduction in whole or in part is permitted for  
any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM												
1. REPORT NUMBER Research Report 77-3	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER												
4. TITLE (and Subtitle) Accuracy of Perceived Test-Item Difficulties		5. TYPE OF REPORT & PERIOD COVERED Technical Report												
		6. PERFORMING ORG. REPORT NUMBER												
7. AUTHOR(s) J. Stephen Prestwood and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0243												
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, Minnesota 55455		10. PROGRAM ELEMENT PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.:61153N PROJ.:RR042-04 T.A.:RR042-04-01 W.U.:NR150-382												
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE May 1977												
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 25												
		15. SECURITY CLASS. (of this report) Unclassified												
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE												
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.														
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)														
18. SUPPLEMENTARY NOTES														
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) <table border="0"> <tr> <td>testing</td> <td>sequential testing</td> <td>programmed testing</td> </tr> <tr> <td>ability testing</td> <td>branched testing</td> <td>response-contingent testing</td> </tr> <tr> <td>computerized testing</td> <td>individualized testing</td> <td>automated testing</td> </tr> <tr> <td>adaptive testing</td> <td>tailored testing</td> <td></td> </tr> </table>			testing	sequential testing	programmed testing	ability testing	branched testing	response-contingent testing	computerized testing	individualized testing	automated testing	adaptive testing	tailored testing	
testing	sequential testing	programmed testing												
ability testing	branched testing	response-contingent testing												
computerized testing	individualized testing	automated testing												
adaptive testing	tailored testing													
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) <p>This study investigated the accuracy with which testees perceive the difficulty of ability-test items. Two 41-item conventional tests of verbal ability were constructed for administration to testees in two ability groups. Testees in both the high- and low-ability groups responded to each multiple-choice item by choosing the correct alternative and then rating the item's difficulty relative to their levels of ability. Least-squares estimates of item difficulty, which were based solely on the difficulty ratings, correlated highly with proportion-correct and latent trait estimates of item</p>														

difficulty based on a norming sample. Least-squares estimates of testee ability, which were based solely on the difficulty perceptions of the testees, correlated significantly with number-correct and maximum-likelihood ability scores based on the testees' conventional responses to the items. These results show that item-difficulty perceptions were highly related to the "objective" indices of item difficulty often used in test construction, and that as testee ability level increased, the items were perceived as being relatively less difficult. The relationship between a testee's ability and his/her perception of an individual item's relative difficulty appeared to be weak. Of major importance was the finding that items which were appropriate in difficulty levels from a psychometric standpoint were perceived by the testees as being too difficult for their ability levels. The effects on testees of tailoring a test such that items are perceived as being uniformly too difficult should be investigated.

## CONTENTS

Introduction .....	1
Method .....	2
Test Construction .....	2
Procedure .....	3
Subjects .....	3
Test administration .....	3
Design .....	4
Accuracy of Difficulty Perceptions .....	4
Method of Analysis .....	4
Difficulty perception model .....	4
Accuracy of ratings-based estimates .....	5
Dimensionality of difficulty perceptions .....	6
Results .....	6
Dimensionality of difficulty perceptions .....	6
Accuracy of ratings-based estimates .....	8
Difficulty Perceptions of Individual Items .....	8
Method of Analysis .....	12
Results .....	12
Perceptions of Appropriate Item Difficulty .....	12
Method of Analysis .....	12
Results .....	13
Discussion .....	13
Conclusions .....	15
References .....	16
Appendix A: Item Calibration Procedure .....	18
Appendix B: Supplementary Tables .....	24

Technical editing by Terry Graham

## ACCURACY OF PERCEIVED TEST-ITEM DIFFICULTIES

Conventional ability tests require all testees to answer the same set of test items. Because testees differ in ability level, however, tests of this kind may potentially create differential psychological environments for testees of different ability levels. A test which is appropriately difficult for a testee of average ability may be perceived by less able individuals as being much too difficult, and such perceptions may lead these testees to approach the task with anxiety and forbearance. On the other hand, individuals with higher than average abilities may find the task a simple or even pleasant one. Clearly, the psychological environment of a testee may vary greatly depending on the individual's perception of the task.

Adaptive tests are designed such that each testee receives items which are psychometrically appropriate for his/her ability level (Lord, 1970; Weiss, 1974; Weiss & Betz, 1973). For example, items in such tests may be chosen so that each testee, regardless of ability level, will have approximately a fifty-percent chance of answering the item correctly (e.g., Lord, 1970). The adaptive test may thus reduce the differential psychological environments arising from the administration of a fixed set of items to persons of differing ability levels, and may thereby improve the performance of low-ability students. In fact, under certain conditions, adaptive testing has been shown to be more motivating for low-ability testees (Betz & Weiss, 1976b) and to result in higher ability estimates (Betz & Weiss, 1976a).

Holtzman (1970) points out the potential importance of psychological factors in the estimation of an individual's ability:

It may be important to investigate the interaction of personality and situational factors with tailored testing. The motivational impact on the student when he discovers that most of the items are at a certain level of difficulty (or uncertainty) is unknown. The optimal level (or mixture of levels) for a given student will not be derived from test theory alone; information about student anxiety and motivation may also be relevant. (p. 199).

Whether adaptive tests can actually reduce the differential psychological effects due to the administration of an inappropriately easy or difficult set of test items depends largely on whether testees can accurately perceive the difficulties of the items administered. Little research has dealt directly with the question of item-difficulty perception.

Munz and Jacobs (1971) asked introductory psychology students to scale multiple-choice examination questions on the subjective difficulty an introductory psychology student would experience in reaching a solution to a particular test question. Thurstone's methods of equal-appearing intervals was used to derive difficulty scale values for the individual items. These scale values correlated positively but moderately ( $r=.52$ ) with traditional proportion-correct difficulty indices based on the subsequent administration of those items to

other introductory psychology students. However, Munz and Jacobs made no attempt to determine the accuracy with which individuals perceived item difficulties relative to their own levels of ability. Further, these results may be generalized only to other achievement-testing situations where students have been exposed to the material and have made an attempt to familiarize themselves with it.

Bratfisch, Dornic, and Borg (1972) asked individuals to estimate the subjective difficulty of items from sets A, B, D, and E of Raven's *Standard Progressive Matrices*. The items were first administered conventionally, in the order of their "objective" difficulty as assessed by determining the proportion of correct responses in a norming sample. Following this, the items were presented in random order and estimates of their subjective difficulties were obtained through a magnitude estimation procedure. The Spearman rank-order correlation between the subjective difficulties of the items and the order of their initial administration (i.e., their ranked "objective" difficulty) was positive and high ( $r_s = .90$ ). Unfortunately, the effect of the items' prior administration in the order of their objective difficulty cannot be determined.

In another study by the same authors (Bratfisch, Borg & Dornic, 1972), testees were administered numerical-reasoning, spatial-ability, or verbal-comprehension items in the order of "objective" difficulty of the items in the tests. Immediately after attempting to answer each item in the conventional manner, the testees rated the item's difficulty on a nine-point scale where 1 corresponded to a "very, very easy" item and 9 corresponded to a "very, very hard" item. The Spearman correlations between order of administration and perceived difficulty for the numerical-reasoning, spatial-ability, and verbal-comprehension tests were .97, .92, and .92, respectively. Unfortunately, in both studies by these authors, the subjective difficulties were not explicitly related to the testees' perceptions of an item's appropriateness to their ability levels. More importantly, in both studies, it is impossible to separate the effect of item difficulty from that of order of administration.

The present study was designed to determine whether or not testees can perceive the difficulties of ability test items relative to their levels of ability and, if so, to investigate the accuracy of these perceptions for individual items. Additionally, the study was designed to determine the level of item difficulty perceived by testees as being appropriate for their ability.

## Method

### Test Construction

Two 41-item conventional tests were designed which had a large range of differences between the difficulties of successive items. Items for the tests were chosen from a pool of five-alternative, multiple-choice vocabulary items on the basis of their normal-ogive difficulty ( $b$ ) and discrimination ( $a$ ) parameters (Lord & Novick, 1968). One of the tests was designed to be administered to a group of relatively low-ability college students. The other test was designed to be administered to a group of relatively higher ability students.



The item parameter estimates were based initially on data reported by McBride and Weiss (1974), derived from samples of University of Minnesota undergraduates. These parameter estimates were revised using a procedure essentially the same as that described by Jensen (1976): Appendix A describes the process of developing the revised item parameters. The difficulty and discrimination parameters for each test item are shown in Appendix Table B-1.

The low- and high-ability tests had a mean difficulty of  $\bar{b} = -2.190$  and  $\bar{b} = -.488$ , respectively. Mean discrimination values for the low- and high-ability tests were  $\bar{a} = 1.117$  and  $\bar{a} = 1.501$ , respectively.

### Procedure

Subjects. Two groups of undergraduate students participated in this study. The first group consisted of 119 students from psychology classes in the University of Minnesota's General College (GC) who were tested in the winter of 1975. The second group, tested in the spring of 1975, consisted of 185 students from an introductory psychology class in the University's College of Liberal Arts (CLA). All students were volunteers who received points toward their final course grades for participation in the experiment. GC students typically perform more poorly on ability and aptitude tests than do CLA students; for the purposes of this study, the GC students will therefore be designated as the "low-ability" group while the CLA students will be referred to as the "high-ability" group.

Test administration. All students were tested at individual cathode-ray terminals (CRTs) connected to a Hewlett-Packard 9600E real-time computer system. Instructional screens similar to those described by DeWitt and Weiss (1974, pp. 36-53) explained the operation of the CRTs before the actual testing was begun. In addition, a proctor was present in the testing room to provide assistance in the operation of the equipment.

Each student answered 41 multiple-choice vocabulary test items. The first six test items presented were identical for testees in a given ability group. These items, whose difficulties reflected the difficulty range of the test, served to familiarize the students with the range of difficulties they would subsequently encounter. The remaining 35 items in each test were presented in four different orders of administration to minimize the effect that the order of item presentation might have on perceived item difficulty. Testees were sequentially assigned to one of the four conditions. Although the same procedure was followed in both ability groups, the items differed between groups. Appendix Table B-1 shows the order of item administration in each of the four conditions for each ability group.

Prior to the administration of the test, the students were informed that they would have as much time as they needed to complete the task. During the test, items were presented on the CRT screen and students responded by typing the number corresponding to the chosen alternative for each five-alternative multiple-choice item. Immediately after responding to an item, each student was asked to indicate the item's perceived difficulty by entering a difficulty code selected from the following list:



- A. Much too easy for you
- B. Somewhat too easy for you
- C. Just about right for you
- D. Somewhat too hard for you
- E. Much too hard for you.

The testee's response was then checked by the computer to ensure that one of the five alternatives had been chosen, and these data were stored with the item-response data for later analysis.

### Design

The study was designed to investigate three different aspects of item-difficulty perception. The initial phase was designed to determine whether or not testees could accurately perceive the difficulty of ability-test items. The second phase was concerned with whether or not a testee's ability level was related to the perception of the relative difficulty of a given item; that is, how accurate an individual's perceptions were, relative to his/her ability level. The third phase of the analysis attempted to determine the relative item difficulty which was perceived by the testee as being about right for his/her ability level.

### Accuracy of Difficulty Perceptions

### Method of Analysis

Difficulty perception model:<sup>1</sup> An individual's perception of an item's difficulty can be thought of as the signed distance between the person's ability level and the item's difficulty level in a Euclidean ability/difficulty space. This perception will be denoted by

$$d_{ij} = \sum_{p=1}^P w_{jp} (x_{jp} - x_{ip}) \quad [1]$$

where  $d_{ij}$  is the perceived difficulty of item  $j$  for person  $i$

$x_{jp}$  is the difficulty of item  $j$  along ability/difficulty dimension  $p$

$x_{ip}$  is the ability of person  $i$  along ability/difficulty dimension  $p$

$w_{jp}$  is the weight of item  $j$  along dimension  $p$

$P$  is the number of dimensions in the ability/difficulty space.

Thus, in this model, the difficulty of an item for a given person is defined as the weighted sum of the signed distances between the location of the item and the location of the person along  $P$  ability/difficulty dimensions. For the present analysis, numerical values of  $d_{ij}$  were assigned to each alternative on

<sup>1</sup>Appreciation for the development of this model is expressed to Mark Davison, Assistant Professor of Educational Psychology, University of Minnesota.

the rating scale. The values assigned to alternatives A through E were -2, -1, 0, +1, and +2, respectively. Thus,  $d_{ij}$  increased as the perceived difficulty of an item increased, and  $d_{ij}$  was equal to zero when an item was perceived by a testee as "just about right for [me]."

The use of a model such as that in Equation 1 is advantageous for several reasons. Using the difficulty ratings alone, estimates of individual ability levels and item difficulties can be derived on a common metric. In addition, the general, multidimensional form of the model may be particularly useful in describing difficulty perceptions on multi-ability test batteries or other such multi-trait instruments.

Note that  $P$  in the model corresponds to the number of dimensions in the space. If the item difficulty ratings are unidimensional,  $P$  will equal 1 and  $d_{ij}$  can be expressed more simply as

$$d_{ij} = w_j(x_j - x_i) \quad [2]$$

Further, if the items are assigned unit weights, the expression in Equation 2 becomes

$$d_{ij} = (x_j - x_i) \quad [3]$$

If the model and the assumption of unidimensionality are appropriate and the average ability level within a group of testees is arbitrarily set at zero, a least squares estimate of a single item's difficulty ( $x_j$ ) is found to be

$$\hat{x}_j = \frac{1}{N} \sum_{i=1}^N d_{ij} \quad [4]$$

where  $N$  is the number of persons rating the item. Thus, an estimate of an item's difficulty is simply the average difficulty rating assigned to that item by the individual being tested.

Similarly, a least squares estimate of  $x_i$ , the ability level of person  $i$ , is

$$\hat{x}_i = -\frac{1}{n} \sum_{j=1}^n d_{ij} + \frac{1}{n} \sum_{j=1}^n \hat{x}_j \quad [5]$$

where  $n$  is the number of items administered. An estimate of an individual's ability level is thus the average difficulty rating he/she assigns to a set of items plus the average item-difficulty in that set.

**Accuracy of ratings-based estimates.** The estimates of item difficulties and individual ability levels described by Equations 4 and 5 are based solely on the testees' ratings of relative item difficulties. In order to determine the appropriateness or accuracy of these perceptions, the ratings-based estimates of item difficulties and students' abilities were compared to more conventional estimates based on the correctness/incorrectness of the testees' conventional responses to the test items.

The ratings-based estimates of item difficulty were correlated with the proportion of persons in the present study identifying the correct response alternative and also with the normal-ogive estimates of item difficulty ( $b_j$ ) based on the item-calibration described in Appendix A. The ratings-based estimates of student ability were correlated with traditional number-correct scores and maximum-likelihood ability estimates (Betz & Weiss, 1976a) based on the normal-ogive parameters of the items.

Dimensionality of difficulty perceptions. In order to use the simple, unidimensional form of the difficulty-perception model described above, the unidimensionality of the difficulty ratings must be demonstrated. Because there is no definitive test of unidimensionality, an indirect evaluation was necessary. McBride and Weiss (1974) suggested four criteria which, if met, constitute sufficient evidence of unidimensionality in item-response data. According to the criteria suggested, confirmatory evidence of unidimensionality is present when: 1) the first common factor of the matrix of inter-item correlations is a general factor accounting for a large proportion of the common variance and on which all variables load highly; 2) the second and subsequent factors account for much smaller and essentially equal proportions of the common variance; 3) the item loadings on the first factor are either all positive or all negative; and 4) none of the above criteria are satisfied by the analysis of a similar correlation matrix constructed from computer-generated random data. Although these criteria were suggested in the context of the analysis of item-response data, they are equally applicable to the analysis of the difficulty ratings.

Accordingly, a  $41 \times 41$  matrix of product-moment inter-item correlations among the difficulty ratings was factor analyzed for each ability group. Communalities for each item were estimated by the squared multiple correlation of that item with all others in the matrix. Factors were extracted by the principal axes procedure and the resulting communalities were substituted for the prior communality estimates. This procedure continued in an iterative fashion until the differences between the two communality estimates were negligible.

## Results

Dimensionality of difficulty perceptions. Evidence of the dimensionality of the difficulty ratings is shown in Figures 1a and 1b. These figures show the first ten eigenvalues of the inter-item correlation matrix based on the difficulty ratings for the low- and high-ability groups, respectively. In both figures, the eigenvalues from the analysis of the ratings are represented by a solid line, while the dashed line shows those resulting from an analysis of comparable, computer-generated random data.

In both ability groups, the first factor of the real data extracted by far the largest amount of variance, while the second factor extracted only slightly more variance than did subsequent factors. The first factors extracted from the random data, on the other hand, accounted for little more variance than other random-data factors. The amount of variance extracted by the second and subsequent factors in the real data was similar to that extracted by the second and subsequent factors in the random data.

Figure 1  
Factor Contributions as a Function of Factor Number for  
the Difficulty Ratings and for Comparable Random Data

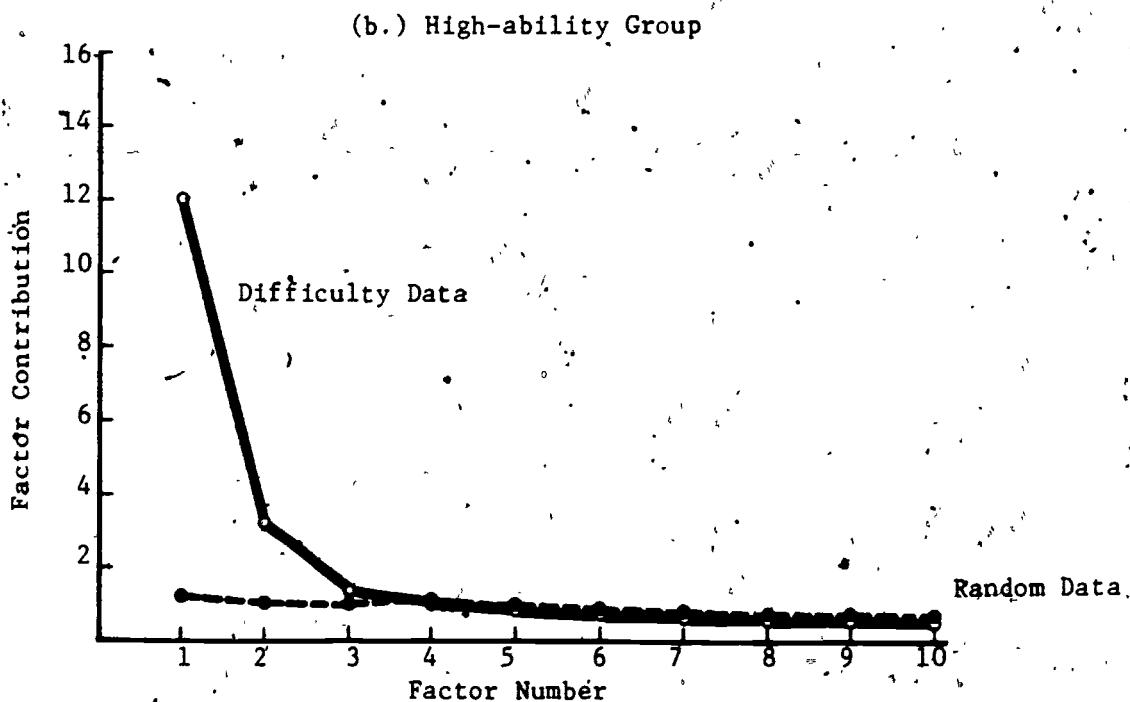
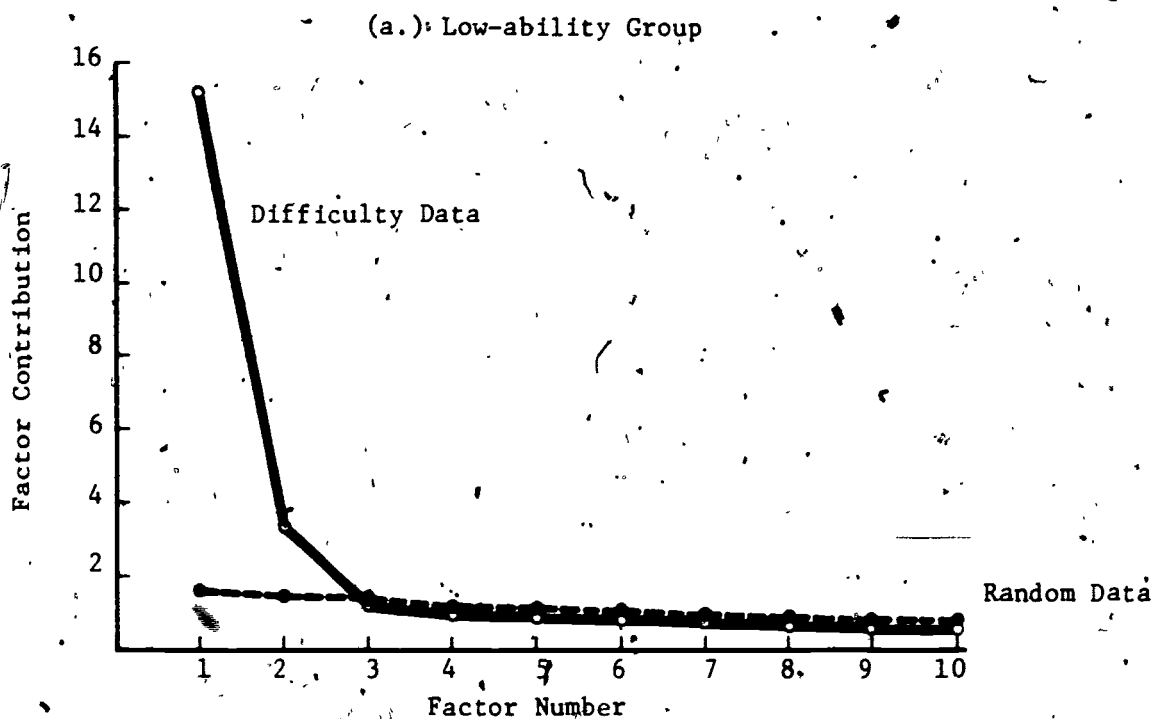


Table 1 lists the loadings of the items from each test on the first three factors extracted from the matrix of inter-item correlations of difficulty ratings for that test. Each of the items loaded positively on the first factor from that test's data, and the first factor loadings were generally high. These data therefore suggest the existence of a "general" factor. Also shown in Table 1 are the loadings for the first three factors from the comparable random data for each group. For these latter data, the first factor was bipolar; i.e., positive and negative loadings occurred as frequently on the first factor as on Factors 2 and 3. In the real data, such bipolarity occurred only on the second and subsequent factors. These results therefore suggest that for both ability groups, the difficulty ratings may be characterized as being unidimensional.

Accuracy of ratings-based estimates. Because the difficulty perceptions appeared to be unidimensional, the difficulty ratings were used in conjunction with Equations 4 and 5 to calculate ratings-based estimates of item difficulty ( $\hat{x}_j$ ) and testee ability ( $\hat{x}_i$ ). The estimates of item difficulties, based solely on the difficulty ratings, are shown in Table 2. Table 2 also shows proportion correct ( $p_j$ ) and normal-ogive ( $b_j$ ) item-difficulty estimates for each item.

In the low-ability group, estimates of item difficulty derived from the difficulty perceptions were highly related to proportion-correct and normal-ogive item-difficulty estimates; Pearson product-moment correlations were  $r = -.86$  and  $r = .80$ , respectively. The relationships between the ratings-based difficulty estimates and the estimates based on conventional responses to the items were similarly high for items in the high-ability group, with respective Pearson product-moment correlations of  $r = -.94$  and  $r = .85$ .

Appendix Table B-2 shows, for each testee, number-correct scores ( $n_i$ ) and maximum likelihood estimates of the testee's ability level ( $\hat{\theta}_i$ ) based on his/her conventional responses to the items and the corresponding ability estimates based on the difficulty perceptions ( $\hat{x}_i$ ). The Pearson product-moment correlations of the ratings-based ability estimates with the corresponding number-correct scores and with maximum likelihood ability estimates were  $r = .55$  and  $r = .56$ , respectively, for testees in the low-ability group. For persons in the high-ability group, comparable correlations were  $r = .63$  and  $r = .59$ , respectively.

#### Difficulty Perceptions of Individual Items

The second phase of the analysis assessed the relationship between the ability levels of testees and the perceived difficulty of a given item. As an individual's ability level increases relative to the difficulty level of an item, the item should be perceived by the individual as being relatively less difficult. As student ability levels decrease in comparison to an item's difficulty, the item should appear to the testees as being relatively more difficult. Thus, the difficulty rating assigned by a testee to an individual item should be dependent upon the discrepancy between the testee's ability level and the item's difficulty.

Table 1  
Item Loadings on the First Three Factors for the  
Difficulty-Perception Data and for Comparable Random Data

Low-Ability Group							High-Ability Group						
Difficulty-Perception Data				Random Data			Difficulty-Perception Data				Random Data		
Item Reference Number	Factor			Factor			Item Reference Number	Factor			Factor		
	1	2	3	1	2	3		1	2	3	1	2	3
2	.73	-.11	-.12	.02	-.11	-.06	2	.62	-.27	.01	.08	-.03	-.05
4	.81	-.34	.07	.43	-.13	-.01	7	.60	-.38	-.16	.32	.07	.25
7	.53	-.28	.06	.13	.20	.29	14	.55	-.43	-.04	.04	.01	.22
14	.73	-.20	.21	.17	.14	-.19	18	.33	-.04	-.26	.12	-.00	-.14
18	.75	-.23	.09	-.36	.05	-.01	19	.47	-.21	.17	-.36	-.09	-.00
19	.55	-.20	-.18	.32	.22	.02	23	.68	-.35	.09	.19	.14	.03
20	.39	-.18	-.01	.25	-.04	.12	24	.57	-.32	.14	-.14	.02	.06
23	.73	-.08	.05	-.01	-.05	.24	39	.50	-.12	-.11	.02	.05	.12
24	.67	-.27	-.10	-.12	.24	.24	44	.61	-.09	.15	-.03	.24	.16
29	.76	-.22	-.01	.15	.32	.16	51	.50	-.03	-.19	.21	.22	.35
41	.76	-.28	.15	.27	-.27	.21	56	.65	.13	.11	.21	.20	-.02
44	.61	.23	-.28	-.27	-.00	-.11	64	.61	-.42	.13	-.35	-.10	-.08
51	.63	.08	.08	.11	.38	-.16	68	.60	-.15	.19	.10	.25	.08
55	.69	-.16	-.09	.14	-.22	-.15	77	.68	-.29	.09	.16	-.30	.10
56	.39	.31	-.19	.15	-.16	.17	86	.57	-.17	-.33	.15	.08	-.30
62	.63	-.27	-.17	-.09	-.13	-.03	91	.58	.12	.01	-.23	-.13	.27
66	.73	-.34	.25	.41	-.18	.12	104	.55	.06	-.25	.30	-.06	.19
68	.61	-.29	.37	-.14	-.02	.14	108	.65	-.21	.03	-.16	-.04	.16
72	.69	-.27	.23	-.14	-.17	.18	111	.51	.27	-.08	-.03	.14	.20
77	.70	.05	-.11	.20	-.13	.07	114	.64	.27	-.14	.07	-.07	.15
78	.74	-.11	-.19	.03	-.16	-.38	115	.42	.25	-.19	.15	-.22	.21
86	.62	.23	-.15	.09	-.23	.15	120	.62	-.10	.04	.19	-.18	.17
89	.67	.17	-.15	.02	.20	.27	137	.49	.24	.03	.04	.25	.01
91	.46	.36	.01	.19	-.10	.32	145	.63	.17	-.21	.11	-.12	-.06
108	.68	.02	-.15	.10	.09	.22	147	.49	.14	.24	.07	-.10	-.23
111	.53	.51	-.16	-.17	.19	.14	154	.55	.08	.04	-.21	.29	-.04
114	.57	.48	.00	.24	-.07	.06	162	.64	.41	-.08	.13	.18	.13
141	.61	.11	.02	-.26	-.12	.17	167	.46	.21	.18	-.19	.04	.02
145	.62	.09	-.07	.09	.21	-.04	174	.43	.09	.11	.09	-.01	.12
154	.59	.36	-.16	.09	.51	.17	182	.66	-.43	.14	.03	.40	-.21
162	.38	.47	.25	-.12	.06	.09	188	.54	.28	.21	-.02	.19	.24
174	.30	.50	.10	-.18	.21	.15	191	.59	-.03	-.10	-.22	.07	.08
182	.71	-.20	.19	-.15	.13	.29	217	.43	.39	.10	-.12	.28	.22
188	.41	.43	.22	.09	.13	.27	253	.17	.34	.34	.03	.13	.04
191	.67	.03	-.37	.26	.04	.11	302	.52	.35	.19	.24	.02	.00
192	.58	.25	-.08	-.01	.12	-.13	319	.49	.35	-.11	-.03	.07	-.00
198	.76	-.08	-.04	.25	-.17	.18	337	.52	.18	.07	.33	-.03	.00
302	.32	.25	.36	-.00	-.05	.25	359	.34	.16	.43	-.02	-.08	-.18
337	.32	.24	.09	-.42	-.29	.32	375	.30	.44	.21	-.18	-.17	-.13
375	.26	.59	.18	.05	.28	-.16	383	.47	.10	-.09	.05	-.09	-.24
651	.39	.46	.31	.12	.12	.02	514	.50	.04	-.42	.12	.29	.05



Table 2  
Least-Squares Item Difficulty Estimates Based on the  
Difficulty Perceptions ( $x_j$ ) and Corresponding Proportion-Correct ( $p_j$ )  
and Normal-Ogive ( $b_j$ ) Item Difficulty Indices

Low-Ability Group				High-Ability Group			
Item Reference Number	$x_j$	$p_j$	$b_j$	Item Reference Number	$x_j$	$p_j$	$b_j$
2	-.38	.99	-3.81	2	-.89	.97	-3.81
4	-.79	.99	-5.56	7	-1.11	.98	-2.32
7	-.96	.97	-2.32	14	-1.10	.96	-2.46
14	-.80	.97	-2.46	18	-.60	.94	-4.24
18	-.59	.94	-4.24	19	-.95	.91	-3.81
19	-.83	.91	-3.81	23	-.97	.99	-3.86
20	-1.51	.96	-5.76	24	-1.15	.99	-2.37
23	-.58	.89	-3.86	39	-.45	.90	-3.63
24	-.83	.99	-2.37	44	-.32	.88	-1.41
29	-.59	.96	-5.52	51	-.09	.75	-1.04
41	-.71	.89	-6.45	56	.39	.47	.13
44	.06	.76	-1.41	64	-1.29	.99	-2.36
51	.18	.57	-1.04	68	-.26	.98	-2.48
55	-.65	.94	-4.95	77	-.75	.94	-3.60
56	.57	.32	.13	86	-.24	.82	-1.19
62	-.94	.99	-4.95	91	-.17	.66	-.20
64	-.97	.97	-2.36	104	.72	.47	.05
68	-.68	.92	-2.48	108	-.36	.75	-1.16
72	-.72	.97	-6.13	111	.72	.34	.94
77	-.29	.79	-3.60	114	.80	.28	.96
78	-.55	.92	-4.84	115	1.23	.16	2.02
86	.21	.59	-1.19	120	-.35	.37	1.46
89	.01	.78	-2.49	137	1.10	.48	-.06
91	.08	.56	-.20	145	.40	.48	.09
108	-.20	.57	-1.16	147	.17	.30	1.47
111	.74	.19	.94	154	.07	.59	-.12
114	.83	.16	.96	162	1.09	.21	1.24
141	-.05	.61	-1.21	167	.67	.41	2.16
145	.16	.47	.09	174	-.84	.30	1.45
154	.22	.42	-.12	182	-1.01	.99	-3.83
162	1.12	.11	1.24	188	-.91	.47	-.04
174	.84	.18	1.45	191	-.30	.89	-1.26
182	-.71	.97	-3.83	217	.97	.28	1.38
188	1.09	.31	-.04	253	1.06	.29	1.44
191	-.16	.76	-1.26	302	.90	.51	.85
192	.36	.89	-6.52	319	1.09	.21	2.14
198	-.51	.94	-2.50	337	.61	.42	1.18
302	.93	.58	.85	359	.59	.16	2.07
337	.76	.41	1.18	375	1.16	.31	.93
375	1.34	.22	.93	383	.94	.34	1.52
651	.84	.31	.89	514	.63	.43	1.74

Table 3  
Correlations of Difficulty Ratings  
with Ability-Level/Item-Difficulty Discrepancy  
( $r$ ) and Dichotomized Item Scores ( $r_{bis}$ )

Low-Ability Group			High-Ability Group		
Item Reference Number	$r$	$r_{bis}$	Item Reference Number	$r$	$r_{bis}$
2	-.39	-.19	2	-.31	-1.00
4	-.27	-.67	7	-.44	-.58
7	-.31	-.30	14	-.33	-.36
14	-.27	-.28	18	-.21	-.60
18	-.34	-.24	19	-.28	-.88
19	-.26	-.78	23	-.38	-.67
20	-.28	-.57	24	-.22	-.07
23	-.37	-.58	39	-.30	-.73
24	-.27	-.30	44	-.25	-.34
29	-.40	-1.00	51	-.39	-.55
41	-.34	-.10	56	-.38	-.75
44	-.49	-.51	64	-.27	.07
51	-.49	-.69	68	-.21	.20
55	-.30	-.30	77	-.36	-.56
56	-.40	-.67	86	-.49	-.66
62	-.26	-.75	91	-.44	-.63
64	-.25	-.15	104	-.44	-.69
68	-.17	.20	108	-.41	-.49
72	-.24	-.73	111	-.38	-.47
77	-.39	-.47	114	-.42	-.41
78	-.56	-.05	115	-.29	-.56
86	-.56	-.66	120	-.31	-.33
89	-.49	-.85	137	-.28	-.61
91	-.34	-.23	145	-.41	-.48
108	-.43	-.40	147	-.13	-.22
111	-.43	-.32	154	-.33	-.38
114	-.43	-.47	162	-.49	-.72
141	-.41	-.48	167	-.23	-.33
145	-.37	-.16	174	-.18	-.18
154	-.51	-.62	182	-.30	.22
162	-.21	-.23	188	-.48	-.65
174	-.23	-.22	191	-.41	-.60
182	-.27	-.30	217	-.23	-.39
188	-.28	-.50	253	.11	-.20
191	-.44	-.52	302	-.31	-.43
192	-.40	-.25	319	-.41	-.61
178	-.35	-.76	337	-.39	-.49
302	-.03	-.37	359	-.01	.14
337	-.19	-.44	375	-.15	-.33
375	-.10	-.06	383	-.36	-.40
651	-.18	-.30	514	-.50	-.45

### Method of Analysis

The normal-ogive testing model permits the estimation of individual ability levels and item difficulty levels on a common metric. Thus, an estimate of the discrepancy between an individual's ability level and an item's difficulty is  $\hat{\theta}_i - b_j$ , where  $\hat{\theta}_i$  represents the ability level of person  $i$ , and  $b_j$  represents the difficulty of item  $j$ .

To assess the relationship between the ability-level/item-difficulty discrepancy ( $\hat{\theta}_i - b_j$ ) and the testee's difficulty perception for a single item ( $d_{ij}$ ), the Pearson product-moment correlation ( $r$ ) between  $\hat{\theta}_i - b_j$  and  $d_{ij}$  was computed for each item. Because the estimate of  $\hat{\theta}_i$  and the estimate of  $b_j$  are fallible and because it is possible that testees' perceptions are more directly related to whether or not they can answer the item correctly than to  $\hat{\theta}_i - b_j$ , the biserial correlation ( $r_{bis}$ ) between the testees' item scores (0 if incorrect, 1 if correct) and their difficulty perceptions was also computed.

### Results

Table 3 shows the correlations of the  $\hat{\theta}_i - b_j$  discrepancy and the difficulty ratings,  $d_{ij}$ , for items on both tests. The median correlations were  $-.34$  for the low-ability group and  $-.33$  for the high-ability group. Correlations ranged from  $-.56$  to  $-.03$  for the low-ability group and from  $-.50$  to  $-.11$  for the high-ability group.

Table 3 also shows the biserial correlations of the item scores and the difficulty ratings for each test item. The median biserial correlations were  $-.40$  and  $-.48$  for the low- and high-ability groups, respectively. These correlations ranged from  $-1.00$  to  $.20$  for the low-ability group and from  $-1.00$  to  $.22$  for the high-ability group.

### Perceptions of Appropriate Item Difficulty

Adaptive testing procedures generally tailor a test such that item difficulty parameters are somewhat near the estimated ability level for a given testee, i.e., so that  $\hat{\theta}_i - b_j$  approaches zero. Although these items may be "about right" in difficulty from a psychometric standpoint, they may not be "about right" from the individual testee's point of view. The third phase of the analysis was designed to determine the testee-ability/item-difficulty discrepancy for an item which was perceived by the testee as being "just about right" for him/her.

### Method of Analysis

For each test item, an average  $\hat{\theta}_i - b_j$  was computed for those persons giving the item rating of "C", indicating that they perceived the difficulty of the item as "just about right" for them.

## Results

Table 4 shows the average  $\hat{\theta}_i - b_j$  discrepancy of subjects assigning "C" to the item for each of the items on the two tests. It is obvious from the data in Table 4 that the "about right" perceptions differ greatly from item to item.

Positive values of these mean discrepancies indicate that an item was perceived as "about right" when the difficulty level of the item ( $b_j$ ) was, on the average, below the testees' estimated ability level ( $\hat{\theta}_i$ ). For the low-ability group, 28 of the 41 items had positive mean discrepancies; these discrepancies ranged from .34 to 5.77. For the high-ability group, 20 of the 41 items had positive mean discrepancies, ranging from .14 to 4.04.

Negative values indicate a judgment of "about right" for items which are above a testee's ability level. For the low-ability group, these ranged from -.31 to -2.04. For the high-ability group, the range was -.06 to -2.44.

The average signed mean discrepancy was 1.358 for the low-ability testees and .2899 for the high-ability testees. These averages are somewhat ambiguous because differing numbers of testees contributed to the computation of means for individual items. The overall mean discrepancies judged to be "about right", weighted by the number of persons upon which each item mean was based, were 1.703 and .466 for the low- and high-ability groups, respectively.

## Discussion

Least squares estimates of item difficulties, based on the difficulty ratings assigned to the items and the unidimensional difficulty-perception model, were closely related to difficulty indices based on conventional responses to the items. Thus, students were able to accurately perceive the relative difficulties of a set of test items. There was some suggestion in the data that high-ability testees perceived item difficulties relatively more accurately than did low-ability testees.

Similarly, ratings-based ability estimates corresponded relatively well with more traditional ability estimates. Because these ratings-based ability estimates were essentially an average of the difficulty ratings assigned to the items, the positive correlations between these estimates and, for instance, the number-correct scores indicate that as ability levels increased, the items were rated as being relatively less difficult, on the average.

The correlations between the ratings-based ability estimates and the number-correct scores also indicate that testees can, with a fair degree of accuracy, perceive how well they have performed on an ability test. The correlations of .55 for the low-ability group suggests that students in this group were slightly less able to perceive their ability levels as assessed by number-correct scores than were testees in the high-ability group, where number-correct scores and ratings-based ability estimates correlated .63. In general, however, the magnitude of the relationships between the difficulty ratings and objective

Table 4  
Mean Signed Discrepancy by Item Between Testee Ability and  
Item Difficulty ( $\hat{\theta}_i - b_j$ ) for Students Rating  
an Item "Just About Right for [me]," for Two Ability Groups

Low-Ability Group			High-Ability Group		
Item Reference Number	Mean Discrepancy	Number of Students	Item Reference Number	Mean Discrepancy	Number of Students
2	2.87	50	2	3.38	60
4	4.63	48	7	1.52	47
7	1.24	36	14	1.68	51
14	1.47	46	18	4.04	58
18	3.37	53	19	3.29	39
19	2.73	42	23	3.16	61
20	4.03	8	24	1.85	43
23	2.97	54	39	3.29	76
24	1.44	46	44	1.15	101
29	4.54	50	51	.79	90
41	5.54	49	56	-.06	59
44	.75	52	64	1.77	34
51	.36	49	68	2.01	82
55	3.94	60	77	2.96	76
56	-.75	35	86	.77	60
62	4.00	38	91	-.29	73
64	1.37	39	104	.14	32
68	1.46	53	108	.85	78
72	5.13	42	111	-.88	48
77	2.66	60	114	-.87	48
78	3.88	62	115	-1.85	11
86	.61	37	120	-1.92	88
89	1.69	51	137	.42	31
91	-.82	53	145	-.26	77
108	.34	54	147	-1.80	84
111	-1.49	32	154	-.15	95
114	-1.25	32	162	-.75	26
141	.50	55	167	-2.44	51
145	-.73	43	174	-1.37	46
154	-.59	63	182	3.16	55
162	-1.45	14	188	.29	32
174	-2.04	26	191	.94	73
182	2.90	61	217	-1.31	38
188	-.31	14	253	-1.99	27
191	.46	49	302	-.96	40
192	5.77	47	319	-1.59	29
198	1.59	57	337	-1.29	63
302	-1.22	20	359	-2.35	42
337	-1.66	35	375	-.62	15
375	-1.37	11	383	-1.20	49
651	-1.59	29	514	-1.64	56
Mean	1.36			.29	
S.D.	2.26			1.84	
Weighted Mean	1.70			.47	
S.D.	2.28			2.05	

estimates of item difficulty and between the ratings and estimates of testees' abilities indicates that testee perceptions of test difficulty and their test performance are, at least generally, accurate.

The second phase of the analysis showed that for an individual item, however, there was relatively little relationship between testee perceptions of item difficulty and testee-ability/item-difficulty discrepancies or the item scores. The median proportions of variance accounted for by the linear relationship between the  $\theta_i - b_i$  discrepancy and the difficulty perceptions ( $r^2$ )

were only .12 and .11 for the two ability groups. The median proportions of variance accounted for by the relationship between the dichotomized item scores and the difficulty perceptions  $r^2_{bis}$  were .16 and .23 for the two groups.

In these latter data, however, there again seems to be a difference in favor of the high-ability group in that their difficulty perceptions were more highly related to their test behavior.

The finding most relevant for the design of ability-testing procedures was that items which were judged by the testees to be "about right" in difficulty were not necessarily "about right" from a psychometric point of view. These data, in fact, show that testees perceived items that were somewhat below their ability levels as being, on the average, about right for persons of their ability level. In the case of the low-ability students, the items perceived as appropriate had, on the average, normal-ogive difficulty parameters which were over 1.5 standard deviations below the testees' maximum likelihood ability estimates. The high-ability students judged items as "about right" if, on the average, they were about one-half standard deviation below their ability levels. Low-ability students tended to judge items as "about right" in difficulty when the items were below their ability levels; the high-ability students divided their "about right" judgements equally between items which were psychometrically too easy and those which were psychometrically too difficult.

### Conclusions

These data show that students' perceptions of the relative difficulties of a set of ability test items are quite accurate, but that their perceptions of the difficulties of individual ability-test items are only moderately accurate. The data also suggest that the ability level of the testee has some effect on difficulty perceptions. Ability level also is related to the accuracy of perception of a testee's own test score. Thus, testees of different ability levels seem to encounter a different psychological environment when interacting with an ability test. This conclusion is further supported by the students' perceptions of the items which are "about right" for their ability levels.

The psychometric and the psychological effects of adapting an ability test to a level where the testee perceives the test difficulty as "about right" should be studied. Adaptive testing strategies usually tailor a test such that the estimated difficulty of each item administered is close to the current estimate of an individual's ability level. In adapting a test to ensure that item difficulties are psychometrically optimal, these strategies may also, in effect, be tailoring the test so that all of the items are perceived by testees as being too difficult for persons of their ability level. The psychological effects of such a procedure should be investigated more fully.



REFERENCES

- Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability test performance. (Research Rep. 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (AD A027147) (a)
- Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing. (Research Rep. 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (AD A027170) (b)
- Bratfisch, O., Borg, G., & Dornic, S. Perceived item-difficulty in three tests of intellectual performance capacity. Stockholm: University of Stockholm, Institute of Applied Psychology, 1972 (29).
- Bratfisch, O., Dornic, S., & Borg, G. Perceived difficulty of items in a test of reasoning ability. Stockholm: University of Stockholm, Institute of Applied Psychology, 1972 (28).
- DeWitt, L. J., & Weiss, D. J. A computer software system for adaptive ability measurement. (Research Rep. 74-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, January 1974. (AD 773961)
- Holtzman, W. H. Individually tailored testing: Discussion. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing and guidance. New York: Harper & Row, 1970.
- Jensema, C. A simple technique for estimating latent trait mental test parameters. Educational and Psychological Measurement, 1976, 36(3), 705-716.
- Lord, F. M. Some test theory for tailored testings. In W. H. Holtzman (Ed.), Computer-assisted instruction, testing and guidance. New York: Harper & Row, 1970.
- Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.
- McBride, J. R., & Weiss, D. J. A word knowledge item pool for adaptive ability measurement (Research Rep. 74-2). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1974. (AD 781894)
- Munz, D. C., & Jacobs, P. D. An evaluation of perceived item-difficulty sequencing in academic testing. British Journal of Educational Psychology, 1971, 44, 195-205.

Urry, V. W. Ancillary estimators for the item parameters of mental test models.  
Paper presented at the Annual Meeting of the American Psychological Association, Washington, DC, August 1975.

Weiss, D. J. Strategies of adaptive ability measurement. (Research Rep. 74-5).  
Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1974. (AD A004270)

Weiss, D. J., & Betz, N. E. Ability measurement: Conventional or adaptive?  
(Research Rep. 73-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, February 1973. (AD 757788)

# APPENDIX A

## Item Calibration Procedures

### Initial Item Parameter Estimates

The item parameterization procedures that were used assumed a normal-ogive latent trait model and the existence of a bivariate-normal joint-distribution of  $\theta$  (levels of the latent ability) and  $x$  (the continuous variable assumed to underlie the dichotomous item responses). Given these assumptions, discrimination ( $a$ ) and difficulty ( $b$ ) parameters may be defined by Equations 6 and 7,

$$a_j = \rho_{\theta x_j} / \sqrt{1 - (\rho_{\theta x_j})^2} \quad [6]$$

$$b_j = \gamma_j / \rho_{\theta x_j} \quad [7]$$

where  $\rho_{\theta x_j}$  is the correlation between individuals' ability levels ( $\theta$ ) and their scores ( $x$ ) on item  $j$ .

$\gamma_j$  is the z-score above which lies the proportion of testees in the population knowing the correct answer to item  $j$  (Lord & Novick, 1968).

In order to estimate  $\rho_{\theta x_j}$ , the biserial correlation ( $r_j$ ) between testees' ability levels and their dichotomized item scores was found by first estimating the point-biserial correlation ( $\hat{r}_j$ ) between ability levels and dichotomous item scores by Equation 8, based on data reported by McBride and Weiss (1974),

$$\hat{r}_j = (\bar{x}_+ - \bar{x}_-) \sqrt{(p_j)(1-p_j)} / s_x \quad [8]$$

where  $\bar{x}_+$  is the mean number-correct score of persons correctly answering item  $j$ ,

$\bar{x}_-$  is the mean number-correct score of persons incorrectly answering item  $j$ ,

$p_j$  is the proportion of persons correctly answering item  $j$ ,

$s_x$  is the standard deviation of number-correct scores for the total group answering item  $j$ .

The biserial coefficient was then computed using the transformation in Equation 9,

$$r_j = \frac{\hat{r}_j \sqrt{(p_j)(1-p_j)}}{\phi[z'_j]} \quad [9]$$

where  $z_j$  is the z-score above which lies the proportion of testees in the norming sample correctly answering item  $j$  ( $p_j$ ).

$\phi[z_j]$  is the density of a normal probability density function at  $z_j$ .

Because a testee could answer an item correctly simply by random guessing on these 5-alternative, multiple-choice items, a guessing parameter ( $c$ ) was defined for each item by Equation 10.

$$c_j = 1/n_j \quad [10]$$

where  $n_j$  is the number of response alternatives on item  $j$ .

In order to account for guessing when the initial  $a$  and  $b$  parameters used to construct the tests described in this report were derived, the estimate of  $p_{0j}$  ( $r_j$ ) computed in Equation 9 was modified according to Equation 11,

$$r'_j = r_j / (1 - c_j). \quad [11]$$

The estimate of  $p_{0j}$  resulting from Equation 11 ( $r'_j$ ) was restricted to the interval from -1.0 to +1.0 and used, along with  $z_j$  (as an estimate of  $y_j$ ), to calculate values of  $a$  and  $b$  for each item using Equations 6 and 7. The resulting values of  $a_j$  were then restricted to the interval from -3.0 to +3.0. The restrictions on  $r'_j$  and  $a_j$  thus affected both the values of the  $a$  and  $b$  parameters but the effects of the restrictions were not necessarily consistent.

#### Revised Item Parameter Estimates

The item parameter estimates derived from the above procedures were used to select items for the tests administered in this study. In the time interval between the construction of the tests and the analysis of the data, it became apparent that certain revisions to these item parameter estimates were necessary for each item. These revised estimates were computed for all 569 items in the pool from which items for this study were selected.

In computing the revised estimates of  $a$  and  $b$  used to analyze the present data, the proportion of testees who actually knew the correct answer to an item ( $p_j$ ) was estimated from the proportion of testees in the population who actually answered the item correctly ( $p_j$ ) and the estimate of  $c_j$ , using Equation 12,

$$p'_j = (p_j - c_j) / (1 - c_j). \quad [12]$$

An estimate of  $\rho_{\theta x_j}$  suggested by Urry (1975) was then computed by Equation 13,

$$z_j^* = \frac{r_j \sqrt{(p_j)(1-p_j)}}{(1-c_j) \phi[z_j^*]} \quad [13]$$

where  $z_j^*$  is the z-score above which lies the proportion of testees in the sample who were estimated to *actually* know the answer to item  $j$  ( $p_j$ ),

$\phi[z_j^*]$  is the density of normal probability density function at  $z_j^*$ .

This estimate of  $\rho_{\theta x_j}$  was then used, along with  $p_j$  as an estimate of  $\gamma_i$ , in Equations 6 and 7 to calculate the revised  $a$  and  $b$  parameters. If  $p_j < c_j$ ,  $p_j$  was set equal to .001. If  $|z_j^*| > .9486833$ ,  $z_j^*$  was set equal to .9486833 with the appropriate sign. This restricted the  $a$ -values to the interval from -3.0 to +3.0 and influenced the  $b$ -values through Equation 7.

This latter procedure<sup>2</sup> differs from that suggested by Jensema (1976) only in that Jensema chose to remove each item from the computation of the test score estimating  $\theta$  during the computation of that item's parameters. For test scores based on large numbers of items, the effects of this exclusion should be negligible.

#### Comparison of Original and Revised Item Parameters

For items in the pool with  $b$  parameters between  $\pm 3.0$ , Figure A-1 presents the bivariate plot of the original and the revised  $b$  parameters. As Figure A-1 shows, the revised  $b$  estimates were closely related to the original  $b$ -values (Pearson product-moment  $r = .98$ ). The bivariate plot of original and revised  $a$ -values is shown in Figure A-2. As this figure shows, the revised  $a$ -values were not as closely related to the original  $a$ -values (Pearson product-moment  $r = .74$ ) as were the revised  $b$ -values.

To determine the effects of the revised item parameters on ability estimates computed using those parameters, maximum likelihood ability estimates were computed using both sets of item parameters for the 185 CLA students involved in this study. The bivariate plot of the two sets of maximum likelihood ability estimates is shown in Figure A-3. The resulting Pearson product-moment correlation of .96 indicated that the ability estimates did not differ greatly depending on whether the original or revised normal-ogive item-parameter estimates were used. This high correlation suggests that essentially the same conclusions would be drawn in this study from the use of either the original set of item parameters or the revised set of parameter estimates based on Urry's (1975) correction procedure.

<sup>2</sup>These procedures were suggested by James B. Simpson of the University of Minnesota.

-21-





Figure A-2  
Joint Distribution of Original and Revised Discrimination  
Parameter ( $\alpha$ ) Estimates

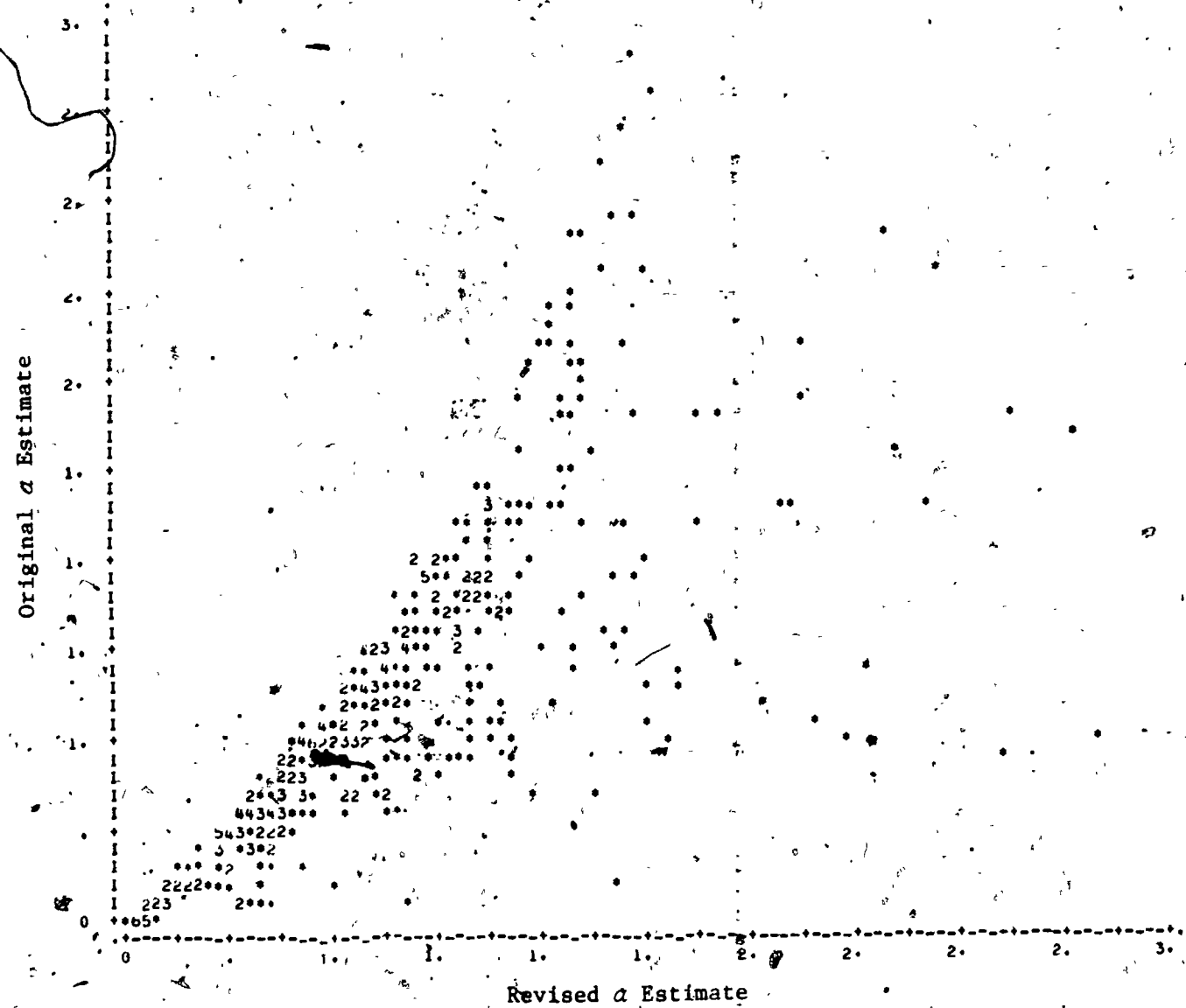
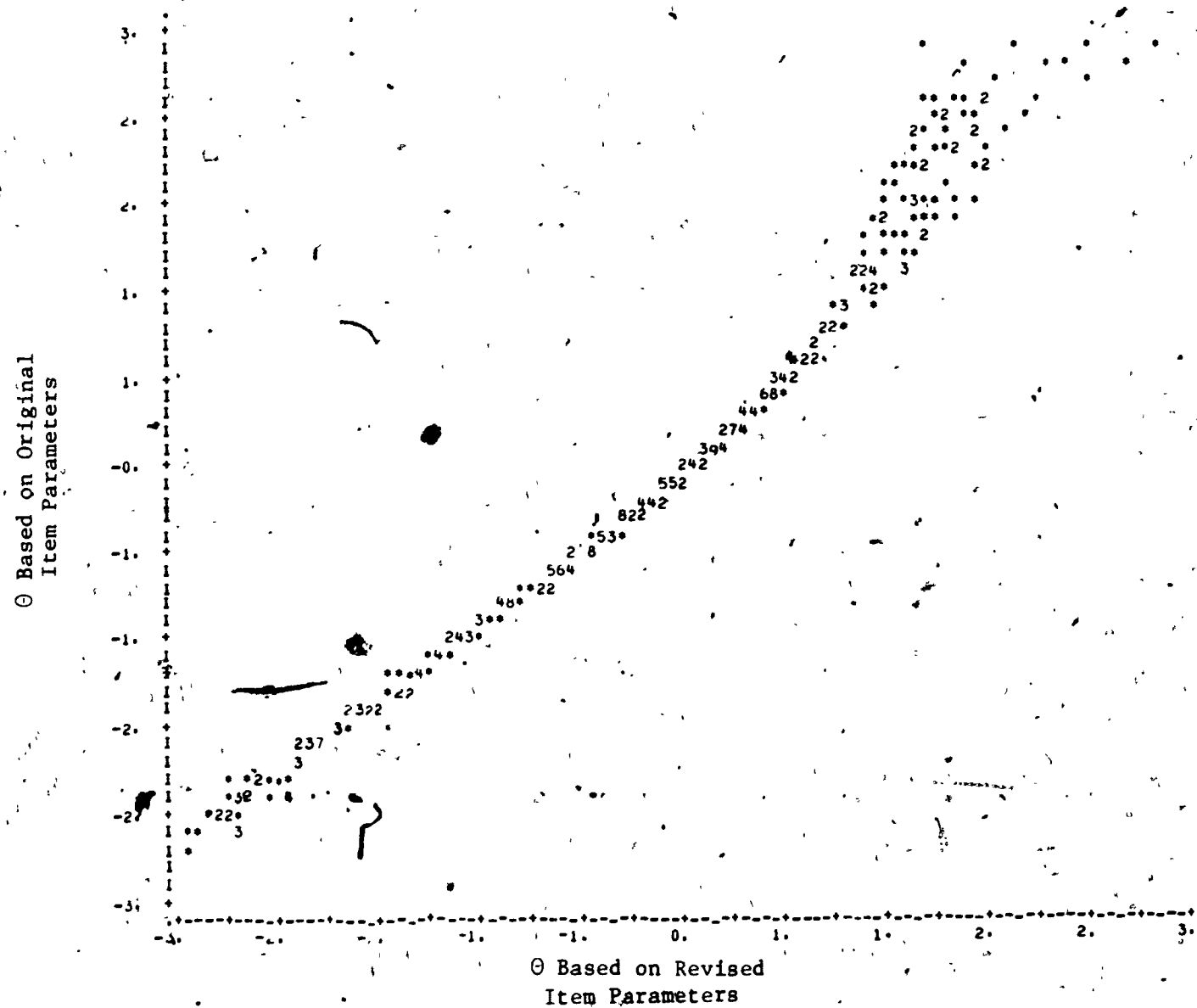


Figure A-3  
Joint Distribution of Maximum-likelihood Ability Estimates ( $\theta$ )  
Based on the Original and the Revised Item-parameter Estimates



APPENDIX B

Table B-1  
Order of Administration and Normal Ogive Discrimination (a) and  
Difficulty (b) Parameters for Items on Tests for the Low- and High-Ability Groups.

Low-Ability Group							High-Ability Group						
Item Reference Number	Item Sequence				Item Parameters		Item Reference Number	Item Sequence				Item Parameters	
	A	B	C	D	a	b		A	B	C	D	a	b
2	11	32	37	16	.517	-3.810	2	41	7	27	21	.517	-3.810
4	24	24	10	38	.397	-5.561	7	39	8	26	22	3.000	-2.324
7	3	3	3	3	3.000	-2.324	14	22	26	8	40	2.208	-2.461
14	40	9	25	23	2.208	-2.461	18	1	1	1	1	.483	-4.241
18	41	7	27	21	.483	-4.241	19	28	20	14	34	.710	-3.808
19	16	37	32	11	.710	-3.808	23	18	39	30	9	.713	-3.862
20	1	1	1	1	.381	-5.764	24	30	18	16	32	1.749	-2.366
23	22	26	8	40	.713	-3.862	39	5	5	5	5	.347	-3.625
24	13	34	35	14	1.749	-2.366	44	32	16	18	30	1.145	-1.412
29	25	23	11	37	.323	-5.521	51	27	21	13	35	1.432	-1.043
41	7	28	41	20	.272	-6.450	56	34	14	20	28	1.109	.135
44	15	36	33	12	1.145	-1.412	64	23	25	9	39	3.000	-2.363
51	34	14	20	28	1.432	-1.043	68	15	36	33	12	1.014	-2.479
55	29	19	15	33	.288	-4.953	77	10	31	38	17	.442	-3.602
56	17	38	31	10	1.109	.135	86	7	28	41	20	.887	-1.189
62	18	39	30	9	.426	-4.952	91	25	23	11	37	1.132	-.197
64	39	8	26	22	3.000	-2.363	104	3	3	3	3	.944	.050
68	6	6	6	6	1.014	-2.479	108	8	29	40	19	.536	-1.155
72	5	5	5	5	.274	-6.134	111	33	15	19	29	.822	.936
77	32	16	18	30	.442	-3.602	114	36	12	22	26	3.000	.960
78	9	30	39	18	.437	-4.843	115	2	2	2	2	3.000	2.023
86	23	25	9	39	.887	-1.189	120	38	10	24	24	3.000	1.464
89	35	13	21	27	.721	-2.493	137	6	6	6	6	.499	-.056
91	30	18	16	32	1.132	-.197	145	35	13	21	27	.791	.086
108	33	15	19	29	.536	-1.155	147	17	38	31	10	.825	1.469
111	19	40	29	8	.822	.936	154	26	22	12	36	.872	-.124
114	8	29	40	19	3.000	.960	162	31	17	17	31	3.000	1.245
141	38	10	24	24	.478	-1.208	167	24	24	10	16	.416	2.155
145	10	31	38	17	.791	.086	174	16	37	32	11	3.000	1.455
154	31	17	17	31	.872	-.124	182	11	32	37	16	.703	-3.833
162	37	11	23	25	3.000	1.245	188	21	27	7	41	.970	-.036
174	36	12	22	26	3.000	1.455	191	37	11	23	25	1.749	-1.257
182	21	27	7	41	.703	-3.833	217	12	33	36	15	1.249	1.384
188	26	22	12	36	.970	-.036	253	14	35	34	13	2.321	1.443
191	12	33	36	15	1.749	-1.257	302	19	40	29	8	.845	.846
192	20	41	28	7	.267	-6.518	319	40	9	25	23	3.000	2.138
198	14	35	34	13	.801	-2.503	337	9	30	39	18	3.000	1.181
302	2	2	2	2	.845	.846	359	29	19	15	33	3.000	2.066
337	4	4	4	4	3.000	1.181	375	20	41	28	7	.832	.934
375	27	21	13	35	.832	.934	383	13	34	35	14	2.111	1.518
651	28	10	14	34	1.087	.885	514	4	3	4	4	1.158	1.741

Table B-2

Least-Squares Estimates of Testee Ability based on the Difficulty Perceptions ( $\hat{x}_i$ )  
with Corresponding Number-Correct Scores ( $x_i$ ) and Maximum Likelihood Ability Estimates ( $\theta_i$ )

Low-Ability Group (N=116)						High-Ability Group (N=185)					
$\hat{x}_i$	$x_i$	$\theta_i$	$\hat{x}_i$	$x_i$	$\theta_i$	$\hat{x}_i$	$x_i$	$\theta_i$	$\hat{x}_i$	$x_i$	$\theta_i$
.30	30	-.12	.08	27	-1.35	.22	28	-.25	-.66	22	-.83
1.00	40	1.80	.08	27	-.96	.36	24	-.56	.34	18	-1.12
-.17	19	-2.13	-.36	19	-2.57	-.34	21	-.37	-.98	12	-1.80
-.26	28	-1.54	1.10	34	.01	-.05	20	-1.92	.41	24	.00
.33	28	-1.20	-.29	28	-.51	-.78	18	-1.03	.46	26	.51
.42	31	-.01	1.37	39	1.41	-.64	20	-.46	-.22	22	-.14
.44	21	-1.93	.08	29	-1.17	-.08	23	-1.05	-.71	21	-1.71
-.24	29	-1.04	.56	27	-.92	.14	29	.43	1.31	31	1.13
-.41	26	-1.43	-.53	29	-.69	-.17	19	-.91	-.25	25	-.00
-.19	31	-.27	-.14	27	-.92	-.15	27	.51	-.22	17	-1.83
1.08	32	-1.04	1.27	34	.00	.39	22	-.68	-.32	24	-.35
-.14	27	-.84	-.41	29	-.95	.63	27	-.01	-.47	18	-1.68
-.51	24	1.79	-.34	27	-.44	-.56	20	-.90	-.27	23	-1.40
.17	30	-.39	1.49	34	.18	.17	17	-1.21	-.20	17	-1.20
-.58	28	-1.17	-.12	24	-1.75	.61	25	-.04	.39	25	.01
-.24	37	-.14	.37	31	-.25	-.32	22	-.76	.70	28	.00
-.14	30	-.53	-.19	28	-1.09	-.12	28	.45	.36	27	.10
.13	25	-1.41	-.17	26	-1.99	-.27	17	-1.86	-1.12	24	-.41
-.14	32	.01	.12	37	.92	-.25	22	-.13	.12	25	-.42
-.31	24	-1.75	.17	32	-.01	-.20	21	-1.06	-.20	21	-.43
-.09	31	-.45	1.13	35	.36	.19	28	-.24	.10	33	1.42
.13	31	-.24	-.41	29	-1.03	-.37	16	-1.55	-.93	18	-1.13
1.03	40	1.47	.10	28	-1.45	.29	30	.20	.78	29	.95
-.39	23	-1.95	-.63	21	-2.31	1.36	37	1.85	.34	26	.12
-.09	31	-.76	-.29	27	-.97	-.39	23	-.63	-.49	22	-.25
1.13	34	.01	-1.44	15	-2.53	.14	18	-2.78	-.20	29	.58
.37	29	-.12	-.26	32	-.01	-.00	35	1.41	-.10	28	.05
-.68	27	-1.08	-.46	29	-1.01	1.83	39	1.96	.56	21	-1.23
-.80	23	-1.93	-.02	23	-1.88	-.27	21	-.68	.95	32	1.20
-.56	28	-.93	-.09	30	-.21	.29	25	-.01	.19	28	.99
-.14	27	-.85	.10	32	-.21	.73	19	-1.27	1.02	39	2.19
-.14	23	-1.78	-.34	29	-.81	-.17	21	-.99	.05	22	-.56
.13	32	-.31	.15	29	-.88	-.61	22	-.59	.46	21	-.32
-.36	28	-1.80	-.02	28	-.80	.34	20	-1.52	.39	27	.14
1.20	34	.01	-.31	33	.01	.38	37	1.90	.46	30	1.17
-.29	29	-.60	-.12	25	-1.27	.34	36	1.77	.34	37	1.61
-.22	30	-.62	1.25	36	.78	-.49	23	-.67	.14	24	-.81
.15	30	-.44	-.70	33	-.83	-.39	23	-.33	-.81	22	-.61
-.75	22	-1.82	.17	28	-.94	-.22	17	-1.66	.07	26	-.41
.74	29	-.22	-1.78	23	-1.85	.22	20	-.78	-.03	19	-1.36
.03	30	-.46	-.29	30	-.33	-.56	18	-1.13	.27	29	.28
.47	29	-.28	.86	33	.12	-.05	29	1.02	-.76	22	-.79
-.29	26	-2.26	1.03	31	-.15	-.25	22	-.71	1.41	34	1.22
-.05	31	-.68	.71	32	-.11	-.32	20	-.80	.14	38	2.28
-.07	25	-1.31	-.24	28	-1.23	1.02	34	1.46	-.00	32	1.37
.74	21	-.31	.13	25	-1.17	.14	30	.17	-.27	26	-.22
-1.12	29	-.37	-.78	24	-1.73	.02	20	-.49	-.37	30	.99
.15	29	-.35	-.75	25	-1.43	-.66	20	-.65	-.73	17	-1.44
.20	26	-1.36	-.05	31	-1.13	.17	29	.48	-.05	19	-.79
.59	31	-.01	.47	25	-1.63	.58	34	1.43	-.20	24	-.41
-.56	21	-2.34	-.17	32	-.00	-.81	10	-2.55	-.59	18	-1.30
-.51	39	1.25	-.46	25	-2.00	-.17	23	-.48	.27	35	1.50
.17	18	-2.28	.98	36	1.00	-.39	19	-.94	.68	35	1.50
-.44	30	-.20	.17	30	-.41	-.20	29	.10	.12	31	1.05
-.36	33	.10	-.24	31	-.20	.56	40	2.49	-.54	19	-.77
-.78	24	-1.70	.27	34	.00	.02	27	-.00	.73	30	1.01
.13	27	-1.14	-.51	32	-.75	-.54	24	.14	.44	35	1.07
-.24	21	-1.97	.08	21	-2.00	-.10	29	.75	-.42	24	-.49
						.58	26	-.23	-.20	23	-.39
						-.54	20	-.45	.41	19	-1.33
						-.56	17	-1.70	.51	24	.12
						.19	35	1.90	1.19	30	.33

# DISTRIBUTION LIST

## Navy

4 Dr. Marshall J. Farr, Director  
Personnel & Training Research Programs  
Office of Navy Research (Code 458)  
Arlington, VA 22217

1 ONR Branch Office  
195 Summer Street  
Boston, MA 02210  
Attn: Dr. James Lester

1 ONR Branch Office  
1030 East Green Street  
Pasadena, CA 91101  
Attn: Dr. Eugene Gloye

1 ONR Branch Office  
536 S. Clark Street  
Chicago, IL 60605  
Attn: Dr. Charles E. Davis

1 Dr. M. A. Bertin, Scientific Director  
Office of Naval Research  
Scientific Liaison Group/Tokyo  
American Embassy  
APO San Francisco 96503

1 Office of Naval Research  
Code 200  
Arlington, VA 22217

6 Commanding Officer  
Naval Research Laboratory  
Code 2627  
Washington, DC 20390

1 LCDR Charles J. Theisen, Jr., MSC, USN  
4024  
Naval Air Development Center  
Warminster, PA 18974

1 Commanding Officer  
U.S. Naval Amphibious School  
Coronado, CA 92035

1 CDR Paul D. Nelson, MSC, USN  
Naval Medical R&D Command (Code 44)  
National Naval Medical Center  
Bethesda, MD 20014

1 Commanding Officer  
Naval Health Research Center  
San Diego, CA 92152  
Attn: Library

1 Chairman, Leadership & Law Dept.  
Div. of Professional Development  
U. S. Naval Academy  
Annapolis, MD 21402

1 Scientific Advisor to the Chief  
of Naval Personnel (Pers Or)  
Naval Bureau of Personnel  
Room 4410, Arlington Annex  
Washington, DC 20370

1 Dr. Jack R. Borsting  
Provost & Academic Dean  
U. S. Naval Postgraduate School  
Monterey, CA 93940

1 Mr. Maurice Callahan  
MODAC (Cbde 2)  
Dept. of the Navy  
Bldg. 2, Washington Navy Yard  
(Acostia)  
Washington, DC 20374

1 Office of Civilian Personnel  
Code 342/02 JAP  
Washington, DC 20390  
Attn: Dr. Richard J. Niehaus

1 Office of Civilian Personnel  
Code 263  
Washington, DC 20390

1 Superintendent (Code 1424)  
Naval Postgraduate School  
Monterey, CA 93940

1 Dr. H. M. West III  
Deputy ADCNO for Civilian Planning  
and Programming (Acting)  
Room 2625, Arlington Annex  
Washington, DC 20370

1 Mr. George N. Graine  
Naval Sea Systems Command  
SEA 047C12  
Washington, DC 20362

1 Chief of Naval Technical Training  
Naval Air Station Memphis (75)  
Millington, TN 38054  
Attn: Dr. Norman J. Kerr

1 Principal Civilian Advisor  
for Education and Training  
Naval Training Command, Code 00A  
Pensacola, FL 32508  
Attn: Dr. William L. Walby

1 Dr. Alfred F. Sade, Director  
Training Analysis & Evaluation Group  
Department of the Navy  
Walter D. FL 32013

1 Chief of Naval Education and  
Training Support (C14)  
Pensacola, FL 32507

1 Naval Anderson Center  
Code 203  
San Diego, CA 92132  
Attn: W. Gary Thomson

1 Navy Personnel R&D Center  
Code 01  
San Diego, CA 92152

5 A. A. Sjonol, Head, Technical Support  
Navy Personnel R&D Center  
Code 201  
San Diego, CA 92152

2 Navy Personnel R&D Center  
Code 410  
San Diego, CA 92152  
Attn: Dr. Martin F. Wiskoff

1 Dr. Robert Morrison  
Navy Personnel R&D Center  
Code 301  
San Diego, CA 92152

1 Navy Personnel R&D Center  
San Diego, CA 92152  
Attn: Library

1 Navy Personnel R&D Center  
San Diego, CA 92152  
Attn: Dr. J. D. Fletcher

1 Officer-in-Charge  
Navy Occupational Development &  
Analysis Center (NODAC)  
Building 150, Washington Navy Yard  
(Acostia)  
Washington, DC 20374

1 Dr. John Ford  
Navy Personnel R&D Center  
San Diego, CA 92152

1 Dr. Worth Scanland  
Chief of Naval Education & Training  
NAS, Pensacola, FL 32508

## Army

1 Technical Director  
U.S. Army Research Institute for the  
Behavioral & Social Sciences  
1300 Wilson Boulevard  
Arlington, VA 22209

1 Armed Forces Staff College  
Norfolk, VA 23511  
Attn: Library

1 Commandant  
U. S. Army Infantry School  
Fort Benning, GA 31905  
Attn: ATSH-IV-IT

1 Commandant  
U. S. Army Institute of Administration  
Fort Benjamin Harrison, IN 46216

1 Dr. Ralph Euser  
U.S. Army Research Institute  
1300 Wilson Boulevard  
Arlington, VA 22209

1 Dr. Leon Lavacki  
U.S. Army Research Institute  
1300 Wilson Boulevard  
Arlington, VA 22209

1 Dr. Joseph Ford  
U.S. Army Research Institute  
1300 Wilson Boulevard  
Arlington, VA 22209

1 Dr. Ralph Carter  
U.S. Army Research Institute  
1300 Wilson Boulevard  
Arlington, VA 22209

1 Dr. James L. Raney  
U.S. Army Research Institute  
1300 Wilson Boulevard  
Arlington, VA 22209

1 Dr. Milton S. Katz, Chief  
Individual Training & Performance  
Evaluation Technical Area  
U.S. Army Research Institute  
1300 Wilson Boulevard  
Arlington, VA 22209

1 HQ USAREUR & 7th Army  
CSCSIPS  
USAREUR Director of GED  
APO New York, 09403

1 ARI Field Unit - Leavenworth  
P. O. Box 3122  
Ft. Leavenworth, KS 66027

1 DCDR, USAADMINTEN  
Bldg. #1, A310  
Attn: AT21-OED Library  
Ft. Benjamin Harrison, IN 46216

#### Air Force

1 Research Branch  
AFHPC/DPMYR  
Randolph AFB, TX 78148

1 AFHRL/AS (Dr. G. A. Eckstrand)  
Wright-Patterson AFB  
Ohio 45433

Dr. Marty Rockway (AFHRL/TT)  
Lowry AFB  
Colorado 80230

1 Instructional Technology Branch  
AFHRL  
Lowry AFB, CO 80230  
ATTN: Major Brian Waters

1 Dr. Alfred R. Fregly  
AFOSR/ML, Building 410  
Bolling AFB, DC 20332

1 Dr. Sylvia R. Mayer (MCIT)  
HQ Electronic Systems Division  
LG Hanscom Field  
Bedford, MA 01730

1 AFHRL/PED  
Stad - 53  
Lackland AFB, TX 78236

1 Major Wayne S. Sellian  
Chief, Personnel Testing  
AFHPC/DPMYR  
Randolph AFB, TX 78148

1 Air University Library  
AUL/LSE 76-443  
Maxwell AFB, AL 36112

#### Marine Corps

1 Director, Office of Manpower  
Utilization  
HQ, Marine Corps (Code MPU)  
USC, Building 2009  
Quantico, VA 22134

1 Dr. A. L. Slafkosky  
Scientific Advisor (Code RD-1)  
HQ, U.S. Marine Corps  
Washington, DC 20380

#### Coast Guard

1 Mr. Joseph J. Cowan, Chief  
Psychological Research Branch (C-P-1/62)  
U.S. Coast Guard Headquarters  
Washington, DC 20590

#### Other DoD

1 Dr. Harold F. O'Neil, Jr.  
Advanced Research Projects Agency  
Cybernetics Technology, Room 623  
1400 Wilson Blvd.  
Arlington, VA 22209

1 Dr. Robert Young  
Advanced Research Projects Agency  
700 Wilson Boulevard  
Arlington, VA 22209

1 Mr. Frederick W. Suffa  
Chief, Recruiting and Retention Evaluation  
Office of the Assistant Secretary of  
Defense, M&RA  
Room 3D270, Pentagon  
Washington, DC 20301

12 Defense Documentation Center  
Cameron Station, Bldg. 5  
Alexandria, VA 22314  
Attn: TC

1 Military Assistant for Human Resources  
Office of the Director of Defense  
Research & Engineering  
Room 3D129, The Pentagon  
Washington, DC 20301

1 Director, Management Information  
Systems Office  
OSD, M&RA  
Room 3B917, the Pentagon  
Washington, DC 20301

#### Other Government

1 Dr. Lorraine D. Eyde  
Personnel R&D Center  
U.S. Civil Service Commission  
1900 E Street NW  
Washington, DC 20415

1 Dr. William Gorham, Director  
Personnel R&D Center  
U.S. Civil Service Commission  
1900 E Street NW  
Washington, DC 20415

1 Dr. Vern Urry  
Personnel R&D Center  
U.S. Civil Service Commission  
1900 E Street NW  
Washington, DC 20415

1 Dr. Andrew R. Molnar  
Science Education Dev. & Res.  
National Science Foundation  
Washington, DC 20550

1 U.S. Civil Service Commission  
Federal Office Building  
Chicago Regional Staff Division  
Regional Psychologist  
230 S. Dearborn Street  
Chicago, IL 60604  
Attn: C. S. Winiewicz

1 Dr. Joseph L. Young, Director  
Memory & Cognitive Processes  
National Science Foundation  
Washington, DC 20550

#### Miscellaneous

1 Dr. John R. Anderson  
Dept. of Psychology  
Yale University  
New Haven, CT 06520

1 Dr. Scarvia B. Anderson  
Educational Testing Service  
Suite 1040  
3445 Peachtree Road NE  
Atlanta, GA 30326

1 Professor Earl A. Alluisi  
Code 287  
Dept. of Psychology  
Old Dominion University  
Norfolk, VA 23508

1 Mr. Samuel Ball  
Educational Testing Service  
Princeton, NJ 08540

1 Dr. Gerald V. Barrett  
University of Akron  
Dept. of Psychology  
Akron, OH 44325

1 Dr. Bernard M. Bass  
University of Rochester  
Graduate School of Management  
Rochester, NY 14627

1 Dr. John Seeley Brown  
Bolt Beranek and Newman, Inc.  
50 Moulton Street  
Cambridge, MA 02138

1 Dr. Ronald P. Carver  
School of Education  
University of Missouri-Kansas City  
5100 Rockhill Road  
Kansas City, MO 64110

1 Century Research Corporation  
4113 Lee Highway  
Arlington, VA 22207

1 Dr. Kenneth E. Clark  
College of Arts & Sciences  
University of Rochester  
River Campus Station  
Rochester, NY 14627

1 Dr. Norman Cliff  
Dept. of Psychology  
University of Southern California  
University Park  
Los Angeles, CA 90007

1 Dr. Allan M. Collins  
Bolt Beranek and Newman, Inc.  
50 Moulton Street  
Cambridge, MA 02138

1 Dr. John J. Collins  
Essex Corporation  
6305 Camino del Estrelo  
San Diego, CA 92120

1 Dr. Rene V. Dawis  
Dept. of Psychology  
University of Minnesota  
Minneapolis, MN 55455

1 Dr. Ruth Day  
Dept. of Psychology  
Yale University  
2 Hillhouse Avenue  
New Haven, CT 06520

1 Dr. Marvin D. Dunnette  
Dept. of Psychology  
University of Minnesota  
Minneapolis, MN 55455

1 ERIC Facility-Acquisitions  
4833 Rugby Avenue  
Bethesda, MD 20814

1 Major I. N. Evonic  
Canadian Forces Personnel  
Applied Research Unit  
1107 Avenue Road  
Toronto, Ontario, CANADA



Dr. Richard L. Ferguson  
The American College Testing Program  
P. O. Box 168  
Iowa City, IA 52240

Dr. Victor Fields  
Dept. of Psychology  
Montgomery College  
Rockville, MD 20850

Dr. Edwin A. Fleishman  
Advanced Research Resources Organization  
8555 Sixteenth Street  
Silver Spring, MD 20910

Dr. John R. Frederiksen  
Polt Beranck & Newman, Inc.  
50 Moulton Street  
Cambridge, MA 02138

Dr. Robert Glaser, Co-Director  
University of Pittsburgh  
3939 O'Hara Street  
Pittsburgh, PA 15213

Dr. M. D. Havron  
Human Sciences Research, Inc.  
7710 Old Spring House Road  
West Cove Industrial Park  
McLean, VA 22101

Dr. Duncan Hansen  
School of Education  
Memphis State University  
Memphis, TN 38118

Human Resources Research Organization  
400 Plaza Bldg.  
Pace Blvd. at Fairfield Drive  
Pensacola, FL 32505

HumRRO/Pestern Division  
27857 Bernick Drive  
Carroll, CA 93924  
Attn: Library

HumRRO/Columbus Office  
Suite 22, 2601 Cross Country Drive  
Columbus, GA 31906

Dr. Lawrence P. Johnson  
Lawrence Johnson & Associates, Inc.  
Suite 502  
2001 S Street NW  
Washington, DC 20009

Dr. Roger A. Kaufman  
203 Good Hall  
Florida State University  
Tallahassee, FL 32306

Dr. Steven W. Keele  
Dept. of Psychology  
University of Oregon  
Eugene, OR 97403

Dr. David Klahr  
Dept. of Psychology  
Carnegie-Mellon University  
Pittsburgh, PA 15213

Dr. Alma E. Lantz  
University of Denver  
Denver Research Institute  
Industrial Economics Division  
Denver, CO 80210

Mrs. W. E. Lassiter  
Data Solutions Corp.  
Suite 211, 6849 Old Dominion Drive  
Falls Church, VA 22101

1 Dr. Frederick M. Lord  
Educational Testing Service  
Princeton, NJ 08540

1 Mr. Brian McNally  
Educational Testing Service  
Princeton, NJ 08540

1 Dr. Ernest J. McCormick  
Department of Psychological Sciences  
Purdue University  
Lafayette, IN 47907

1 Dr. Robert R. MacRie  
Human Factors Research, Inc.  
6780 Corton Drive  
Santa Barbara Research Park  
Goleta, CA 93017

1 Dr. William C. Mann  
University of So. California  
Information Sciences Institute  
4676 Admiralty Way  
Marina Del Rey, CA 90291

1 Mr. Edward Marks  
304 Grange Bldg.  
Pennsylvania State University  
University Park, PA 16802

1 Dr. Leo Munday  
Houghton Mifflin Co.  
P. O. Box 1990  
Iowa City, IA 52240

1 Richard T. Mowday  
College of Business Administration  
University of Nebraska, Lincoln  
Lincoln, NE 68582

1 Dr. Donald A. Norman  
Dept. of Psychology C-009  
University of California, San Diego  
La Jolla, CA 92093

1 Mr. Luigi Petruccio  
2431 N. Edgewood Street  
Arlington, VA 22207

1 Dr. Lyman W. Porter, Dean  
Graduate School of Administration  
University of California  
Irvine, CA 92717

1 Dr. Diane M. Ramsey-Klee  
R-K Research & System Design  
3947 Ridgmont Drive  
Malibu, CA 90265

1 R. Dir. H. Pauch  
P 11 4  
Bundesministerium der Verteidigung  
Postfach 161  
53 Bonn 1, GERMANY

1 Dr. Joseph W. Rigney  
University of So. California  
Behavioral Technology Laboratories  
3717 South Grand  
Los Angeles, CA 90007

1 Dr. Andrew M. Rose  
American Institutes for Research  
1055 Thomas Jefferson St. NW  
Washington, DC 20007

1 Dr. Leonard L. Rosenbaum, Chairman  
Dept. of Psychology  
Montgomery College  
Rockville, MD 20850

1 Dr. Benjamin Schneider  
Dept. of Psychology  
University of Maryland  
College Park, MD 20742

1 Dr. Mark Reckase  
University of Missouri-Columbia  
Dept. of Educ. Psychology  
12 Hill Hall  
Columbia, MO 65201

1 Dr. Robert J. Seidel  
Instructional Technology Group,  
HumRRO  
300 N. Washington St.  
Alexandria, VA 22314

1 Dr. Richard Snow  
Stanford University  
School of Education  
Stanford, CA 94305

1 Mr. Dennis S. Sullivan  
c/o Canyon Research Group, Inc.  
32107 Lindero Canyon Road  
Westlake Village, CA 91360

1 Dr. Keith Wescourt  
Dept. of Psychology  
Stanford University  
Stanford, CA 94305

1 Dr. Anita West  
Denver Research Institute  
University of Denver  
Denver, CO 80201

1 Dr. Earl Hunt  
Dept. of Psychology  
University of Washington  
Seattle, WA 98105

1 Dr. Thomas G. Sticht  
Assoc. Director, Basic Skills  
National Institute of Education  
1200 19th Street NW  
Washington, DC 20208

Previous Reports in this Series

- 73-1. Weiss, D.J. & Betz, N.E. Ability Measurement: Conventional or Adaptive? February 1973. (AD 757788).
- 73-2. Bejar, I.I. & Weiss, D.J. Comparison of Four Empirical Item Scoring Procedures. August 1973.
- 73-3. Weiss, D.J. The Stratified Adaptive Computerized Ability Test. September 1973. (AD 768376).
- 73-4. Betz, N.E. & Weiss, D.J. An Empirical Study of Computer-Administered Two-Stage Ability Testing. October 1973. (AD 768993).
- 74-1. DeWitt, L.J. & Weiss, D.J. A Computer Software System for Adaptive Ability Measurement. January 1974. (AD 773961).
- 74-2. McBride, J.R. & Weiss, D.J. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974. (AD 781894).
- 74-3. Larkin, K.C. & Weiss, D.J. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974. (AD 783553).
- 74-4. Betz, N.E. & Weiss, D.J. Simulation Studies of Two-Stage Ability Testing. October 1974. (AD A001230).
- 74-5. Weiss, D.J. Strategies of Adaptive Ability Measurement. December 1974. (AD A004270).
- 75-1. Larkin, K.C. & Weiss, D.J. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975. (AD A006733).
- 75-2. McBride, J.R. & Weiss, D.J. TETREST: A FORTRAN IV Program for Calculating Tetrahoric Correlations. March 1975. (AD A007572).
- 75-3. Betz, N.E. & Weiss, D.J. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975. (AD A013185).
- 75-4. Vale, C.D. & Weiss, D.J. A Study of Computer-Administered Stradaptive Ability Testing. October 1975. (AD A018758).
- 75-5. Weiss, D.J. (Ed.). Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975. (AD A018675).
- 75-6. Vale, C.D. & Weiss, D.J. A Simulation Study of Stradaptive Ability Testing. December 1975. (AD A020961).
- 76-1. McBride, J.R. & Weiss, D.J. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976. (AD A022964).
- 76-2. Miller, T.W. & Weiss, D.J. Effects of Time Limits on Test-Taking Behavior. April 1976. (AD A024422).
- 76-3. Betz, N.E. & Weiss, D.J. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976. (AD A027147).
- 76-4. Betz, N.E. & Weiss, D.J. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976. (AD A027170).
- 76-5. Pine, S.M. & Weiss, D.J. Effects of Item Characteristics on Test Fairness. December 1976. (AD A035393).
- Weiss, D.J. Final Report: Computerized Ability Testing, 1972-1975. April 1976. (AD A024516).
- 77-1. Weiss, D.J. (Ed.). Applications of Computerized Adaptive Testing. March 1977. (AD A038114).
- 77-2. Vale, C.D. & Weiss, D.J. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.

AD Numbers are those assigned by the Defense Documentation Center, for retrieval through the National Technical Information Service.

Copies of these reports are available, while supplies last, from:

Psychometric Methods Program  
Department of Psychology  
University of Minnesota  
75 East River Road  
Minneapolis, Minnesota 55455